一种快速的天文图像模拟方法及分析*

张鑫[1,2]
(1. 中国科学院国家天文台空间天文与技术重点实验室，北京，100101； 2. 中国科学院大学，北京，100049)

摘 要：图像模拟在天文学研究中已经发挥越来越重要的作用，其重要性主要表现在以下几个方面：通过模拟可以为准备研制的天文观测设备评估提供重要的依据；通过对模拟数据的处理可以验证数据处理的算法。本文对图像模拟的整个过程进行描述，并在计算上对图像模拟的局部方法进行优化，使模拟结果能够高效准确的生成。本文所做的主要工作包含以下几个方面：对天体的形态的模拟，包括点源（如恒星）和展源（星系），而这里主要是对展源建模；模拟观测条件，主要是仪器和天光背景等产生的噪声；模拟点扩散函数（PSF）；从计算和程序上的方法对计算流程进行优化；对结果进行分析。本文中所采用的数据是哈勃超深空场（HUDF）ACS WFC i(F775)波段的数据，其极限星等达到

关键词：图像模拟； HUDF； 星系

引 言

天文图像模拟是将图像处理的技术与天文学模型相结合，这种模拟在天文学研究中已经越来越广泛的使用。模拟有遇见未来之功效，通过模拟可以对未来的设备进行更加科学的评估。基于这种原因，越来越多人开始从事天文图像模拟方面的研究。在对天文的观测研究中，有一些被学者和观测者熟识的天文观测模拟软件，如 WorldWide Telescope、Astronomy CCD Calculator、Stellarium 等，这些软件对天体空间位置、天体特性描述的非常细致，给观测者带来了很大的便利，但是对于遥远星系的描述这些软件还不是很完善。随着大型观测设备的出现，人们对未知天空的观测变的越来越遥远，星系模拟也成为评估大型天文设备很重要的一部分。Bertin[1]在其文中介绍了 SkyMaker，该软件很细致的描述了点扩散函数，包括了大气扰动、望远镜运动、像差、衍射等各个因素，对于星系的形态其利用了 De Vaucouleurs 形态。Refregier[2]等人利用 shapelet 的方法对星系的形态进行模拟，该方法对星系形态的模拟采用多维的模拟方法，文中与 wavelet 等方法进行比较，对残差进行分析，证明其方法的可靠性。LSST[3]成立了专门的图像模拟团队，该团队主要的工作就是为 LSST 做模拟工作。他们的模拟工作是针对全天区的模拟，利用宇宙模型建立星表来确定天体的分布，并且建立了多种天体现象的模型（天体模型包括恒星、星系、小行星等），在传输过程中考虑天气情况、大气扰动等，最后还考虑了结合巡天设备影响所产生的 PSF 函数。利用准确的模拟结果可以对 LSST 系统的各项性能进行分析处理。另外，LSST 对天体形态的模拟是采用了蒙特卡罗的方法，模拟光子的射入。Joel Bergé[4] 等发表的文章中提到了模拟的两种不同方法，一种是类似 LSST 模拟光子射入的方法，另一种是按照像素模拟的方法，并比较两种方法的差异，该文章用到了简单的数学模型，主要强调了处理的迅速。

*基金项目：国家自然科学基金资助项目(U1731127)资助.
收稿日期：； 修订日期：
作者简介：张鑫，男，博士，研究方向：天文数据模拟与数据处理。 Email:zhangx@nao.cas.cn

本文中所做的模拟是利用 Hubble 望远镜的数据和观测条件。文中的图像模拟所做的工作包含：1）模拟天体分布和天体的形态，主要是指星系的形态；2）模拟观测条件以及各种噪声，包括天光背景、暗流、读出噪声等；3）模拟仪器的 PSF；4）对观测结果进行统计分析。图 1 为本文图像模拟的流程图。

本文的主要目标是利用简单通用的模型对天文图像实现快速的模拟。在程序编写上应用更多的优化，使程序达到快速可用的目的。本文还将利用 SExtractor 对模拟结果进行提取，将提取的结果与参考结果进行比较分析，通过分析确定模拟结果在实际应用上的可行性。本文 1 部分介绍图像模拟的几个关键模型和方法，主要包括星系形态的模拟模型、PSF 模型以及观测条件模拟等；2 部分是对模拟结果的分析；3 部分对整个方法流程总结。

图 1 图像模拟流程图
Fig. 1 Flow Diagram of Image Simulation

1 图像模拟的关键模型及方法

本文中的模拟所利用的数据是 HUDF ACS WFC i（F775）波段的数据，该数据生成的图像约为 10000*10000 个像素，由于图像在空中存在一定的旋转，图像四个角包含未拍摄的区域，本文为了实验方便，裁取了其中 4500*4500 个像素的中间区域，面积约为 5.1 $arcmin^2$。文中的模型采用了天文学中一些简单高效的模型，主要目的就是为了对数据进行快速处理。在观测条件上，模拟了 Hubble 望远镜在 600KM 的高空中的观测条件。

1.1 星系的模型

文中的星系的模型采用了 Sersic 模型 [5]，数学表示如下：

(1)

式(1)中表示 R 处的流量密度，表示有效半径处的流量密度，为有效半径，n 为 Sersic 系数，通过该系数可以控制星系的弥散程度，为常数。

通过变换可以将(1)式子变换如下表示方法：

(2)

(2)式中表示中心位置的流量密度，n 为 Sersic 系数，a 为常数，表示当 R 为 a 时的流量密度为中心位置流量密度的

。

若星系为圆形区域，对(2)式在圆面内积分，得到星系的流量

(3)

(3)式中为不完全 gamma 函数，。当积分的半径趋于无穷时

(4)

(4)式中为 gamma 函数。

在模拟中，通常用椭率来表示星系的形状，这就需要将星系的流量密度分布在

整个椭圆平面内，定义椭圆面的椭率为 e，则(2)式中 R 可以用归一化的椭圆半径表示

(5)

(5)式中表示坐标为，坐标的原点为椭圆的中心。

星系在图像中的表现，除了弥散程度和形状的特性外，还存在着方向性，这就需要对已经计算出来的椭圆形的 Sersic 分布进行坐标变换，假设星系坐标系（长轴为 x 轴，短轴为 y 轴）与图像坐标系的夹角为，则变换矩阵为

则新的坐标位置为

(6)

通过(1)-(6)的计算可以模拟出星系的形状、大小、方向等要素，而对星系的模拟所需要的参数，如（Sersic 系数）、（有效半径）、亮度等参数都是通过对 HUDF ACS WFC 的数据进行提取获得。在 Coe D.等[6]的文中利用 SExtractor 对 HUDF 的数据进行提取，并和 Beckwith[7]提供的数据进行对比，结果比较理想，文中所有用到的参数均来自于 Coe D.等所提取的星表。

1. 2 观测条件及噪声的模拟

对于观测条件的模拟，在图像中表现其实就是对噪声的还原，还原的真实与否是影响模拟图像效果的最主要因素。在观测中所产生的噪声主要包括固有噪声、信号噪声。

固有噪声是指源于探测器自身的噪声，包括热噪声、散粒噪声、生成-复合噪声和闪变噪声，在文中用读出噪声将其表示。为了和 Hubble 的数据进行对比，读出噪声是利用 Hubble 的曝光时间计算器[8]计算得出。

信号噪声在文中主要包含的是背景噪声和由于信号的量子性所产生的随机噪声。背景噪声利用了 HST 测得的天光背景噪声，包括黄道光和赤道光。

对于信号，由于其量子性，到达探测器会有一定的随机性，基于这种特性也会产生一定的噪声，在文中用泊松分布表示这种特性。

1. 3 点扩散函数模拟

对于地面望远镜，影响点扩散函数的主要因素有大气扰动和仪器自身的光学特性。而对于 Hubble 望远镜这样的空间设备来说，观测条件要远远好于地面望远镜，因此对于点扩散函数的影响仅仅是来自观测设备的光学特性。

文中所做的模拟是对空间望远镜成像模拟，点扩散函数采用 Gaussian 函数表示，如下：

(7)

上式中，I_0为常数，表示中心位置的流量密度，可以控制 Gaussian 函数的幅度，[x, y]为图像坐标，σ 用来控制图像的质量，对于 Gaussian 函数通常可以通过所给的半高全宽（FWHM）所求得。

(8)

而对于地面的望远镜通常需要考虑大气扰动的效应，所以模拟中所用到的点扩散函数需要考虑观测设备和大气扰动的综合效应，对于这类点扩散函数模拟通常

采用Moffat[9]模型,如下：

(9)

其中，I_0表示中心位置处的流量密度，α 和 β 为尺度因子，可以控制函数的弥散程度，[x,y]为图像坐标，α 和 β 关系如下：

(10)

文[10]中提到，当 β=4.765 时，Moffat PSF 可以很好的模拟大气扰动。

1.4 子像素级处理

模拟的其中一个目的就是对未来某一个观测设备的评估，文中所模拟所利用的数据只符合 Hubble ACS 的特性，例如像素的大小为 0.03×0.03 $arcsec^2$(Drizzling 之后的像素大小)，为了使程序能够为其他观测设备评估所利用，文中的模拟需要做到子像素级。至于所要做到的子像素级别需要根据具体的情况而定。

本文所做的子像素级处理主要是应用在天体形态、亮度的模拟过程中，根据需要将原像素分成若干个子像素（若干个子像素能将原像素均分），根据天体的特性，在子像素中计算得到该像素的值。这一计算，单一看来十分简单，但综合起来，计算量会变得十分庞大，这就需要对程序进行其他优化。

1.5 并行计算

在计算天体的形态、亮度的要素时，由于将原有像素拆分为子像素计算，这样会导致计算量倍增。另外，每一个天体都是独立的个体，单独计算对其他天体并不会产生任何影响。所以，文中将计算每一个天体的形态、亮度等要素进行独立计算，这样，将模拟中的所有天体按照 CPU 的数目进行平均分配，不同的 CPU 计算完之后的结果返给主进程进行合并，形成最终模拟的整幅图像。

2 结果分析

2.1 参考样本分析

文中用到 Coe D. 等提供的数据，该数据中包含了提取出的 Sersic 系数，而 Beckwith 提供了一份更加准确的数据，不过在该数据中对于天体弥散程度的描述只给出了 Gaussian 的参数，但是其位置和亮度要更加准确，事实上，Coe D. 等提供的数据中也给出了一些与 Bechwith 的数据的对比，例如，前者的数据中给出了中心位置与 Bechwith 的对比，本文对两个数据做了更详细的对比。

图 2 是对两份数据提取天体的亮度的对比，蓝色的线为 Coe D. 等的数据天体亮度的统计，红色的线为 Bechwith 的数据天体亮度的统计。从该图中可以看出，两份数据的亮度存在着一定偏差，Coe D. 等提供的数据天体亮度要比实际偏大。

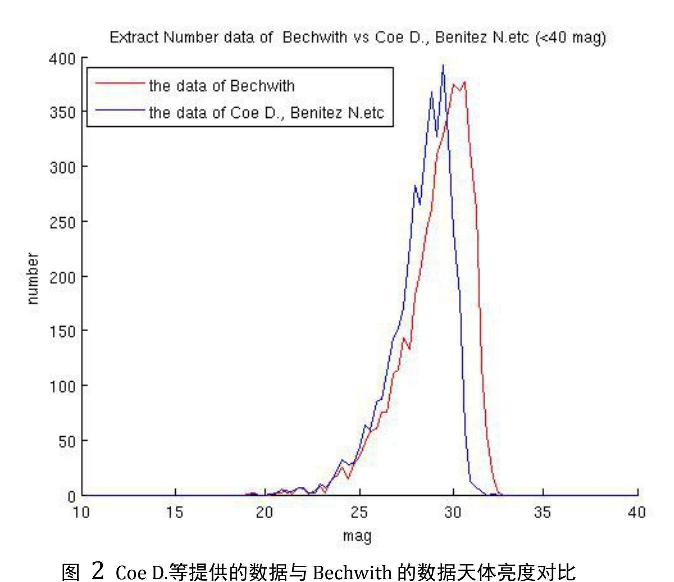


图 2 Coe D.等提供的数据与 Bechwith 的数据天体亮度对比

Fig. 2 AB magnitude comparison between Coe et al. catalog and Bechwith et al. catalog

图 3 是对两个样本有效半径的对比结果，横坐标是提取天体的亮度信息，纵坐标是 Benchwich 的数据中天体的有效半径与 Coe D. 等的数据中的相应的天体的有效半径之差。从该图中可以看出，差值绝大部分都集中在 0 值附近，但是 Coe D. 等的数据有部分半径会偏大。

综合上述对样本的对比，我们利用 Coe D. 等人提供的数据所模拟的图像与真实图像会存在亮度偏差，另外，对于天体的大小也会有部分天体与真实图像存在一定差别。从这一点，也可以解释后面章节中实验结果图 4 的模拟结果存在一些比原图中大的天体的现象。

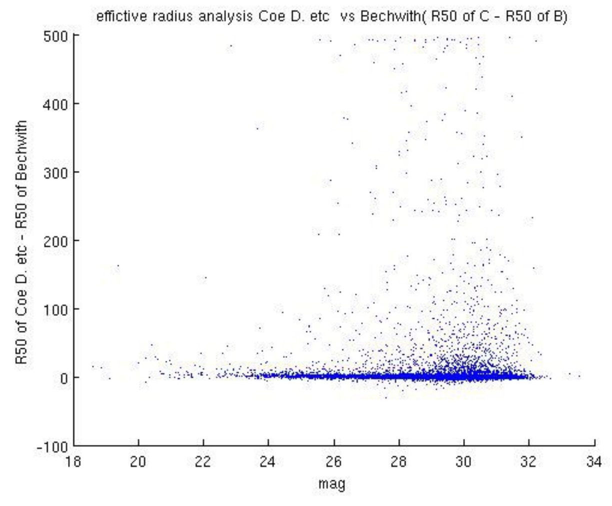


图 3 两种参考样本提取天体有效半径的对比

Fig. 3 Effective Radius comparison between Coe et al. catalog and Bechwith et al. catalog

2.2 模拟结果及分析

图 4 展示了模拟的结果，a 图是模拟图像，b 图是 drizzling 之后原始图像。由于原图在空间中存在一定角度，并且四个角上没有目标信息，所以本文所用到

的数据和图像是对原图的截取，截取了原图 x 方向 3001 到 7500 像素，y 方向 3001 到 7500 像素的方形区域，总共 4500*4500 个像素，drizzling 之后每个像素的视场约为 0.03×0.03 $arcsec^2$，模拟区域的视场为 5.1 $arcmin^2$。HST 团队给出 ACS WFC i（F775）波段零点星等（zeropoint）的值为 25.654，这里零点的意思就是流量为 1 e-时天体的亮度值。根据式子(11)可以计算出天体的流量

(11)

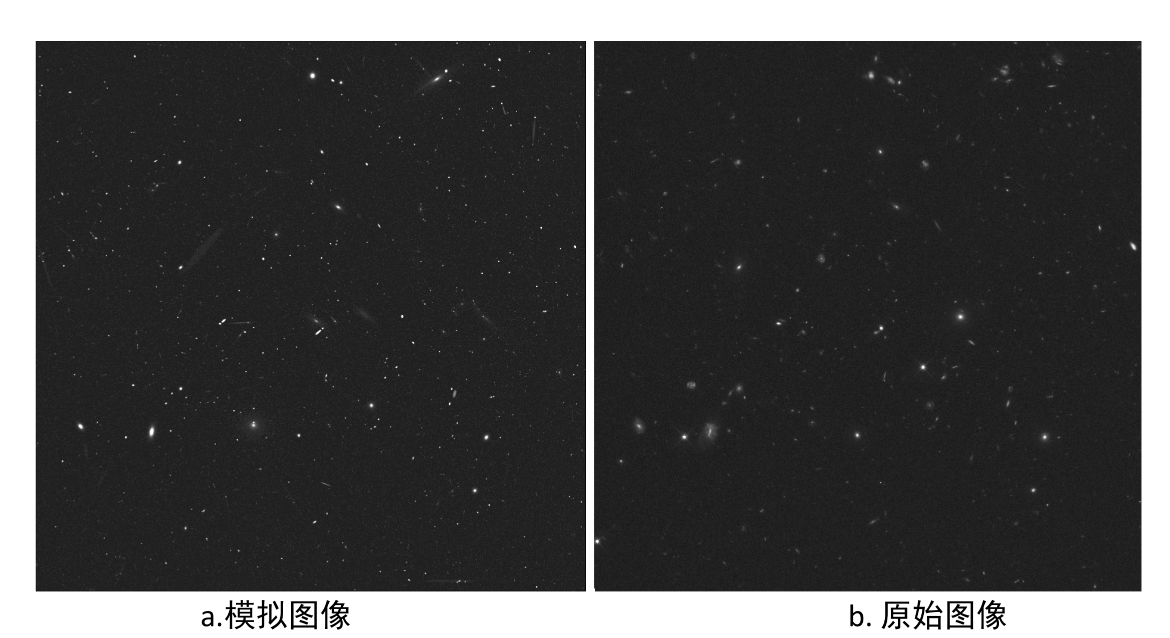
由式(11)所计算出的流量值是信号的流量。而噪声包括天光背景、暗流、读出噪声等，是利用 hubble 提供的曝光时间计算器计算得到。计算的结果为以 e- 为单位的流量。所有计算的结果除以增益（单位，e-/DN）就得到表现在图像上的数字值。

HUDF ACS WFC I 波段的极限星等为 29 等，所以本文提取的截止亮度到 29 等。文中用 SExtractor 作为提取工具，SExtractor 的提取主要参数见附件 1。表 1 是对提取数目的统计，参考的对象为 Coe D.， Benitez N. 等提取的数据。表中有四项数据，第一项是对原图提取数据的统计，第二项是对模拟图像提取数目的统计，第三项是粗略的做了匹配统计，匹配的方法是将模拟数据提取出来的天体的中心与 Coe D.， Benitez N. 等提取的数据的天体的中心作比较，选出距离最接近的一对被认为是同一个天体。第四项是计算匹配率的比例，该比例通过式(12)计算得到

(12)

上式中 ref num 是参考数据提取天体的数目，matched sim num 是对模拟数据提取天体的数目。从表 1 中匹配的数目可以看出，这种针对位置的匹配度均在 85%以上，考虑到噪声影响和提取方法的差异，这种匹配程度是可以接受的。

图 5 是对小于 29 等的天体数目进行统计，从该图中可以看到，对于模拟图像所提取的天体数目和参考数据天体的数目随着亮度变化的趋势是比较一致的。



a.模拟图像　　b.原始图像

图 4 模拟图像结果

Fig. 4 Simulation Image vs. Original Image(a: Simulation Image, b: Original Image)

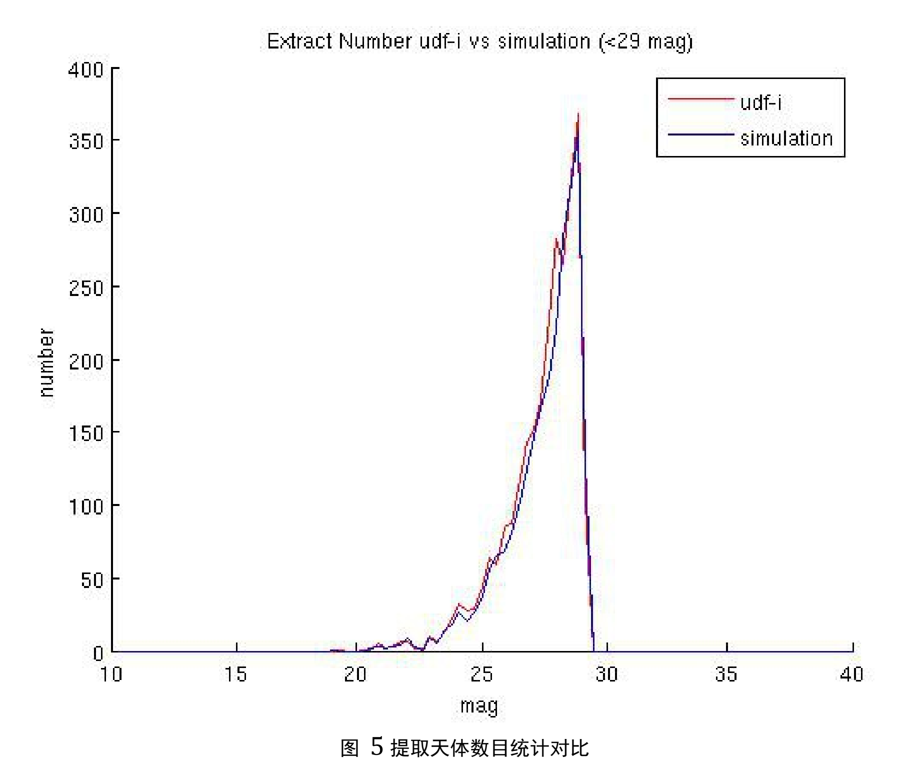


图 5 提取天体数目统计对比

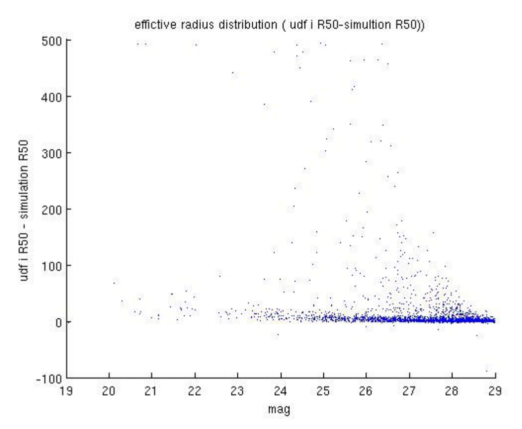
Fig. 5 Extracted Galaxies from Simulation Image vs. HUDF catalog

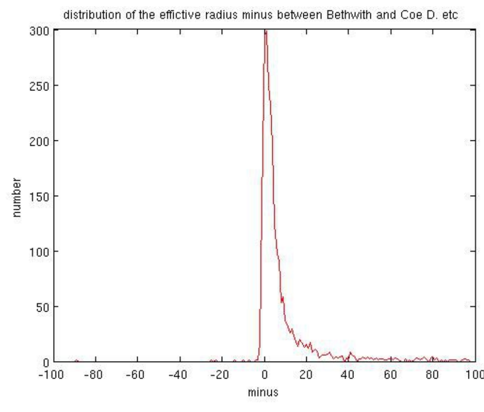
表 1 利用模拟图像提取的天体与 HUDF catalog 天体匹配数目统计

Tab. 1 Extracted galaxies analysisi

	参考数据中天体数目	根据模拟提取的天体数据数目	匹配的数目	匹配率
<26 mag 匹配天体数目	452	406	391	86.5%
<27 mag 匹配天体数目	871	772	753	86.45%
<28 mag 匹配天体数目	1588	1400	1363	85.83%
<29 mag 匹配天体数目	2655	2490	2391	90.1%

图 6 是将模拟数据提取出的天体的有效半径与参考数据天体的有效半径进行对比，采用了计算匹配天体有效半径之差的方法进行对比，匹配天体的方法见表 1 说明，在这里只是确定了位置上的匹配，这两个统计的样本都是亮度小样 29 等天体。图 6（a）统计该差值随着亮度分布的变化，横坐标是亮度，纵坐标是参考数据中的天体的有效半径与相应的模拟数据提取的天体的有效半径之差，从该分布图上可以看出，亮度越小，其有效半径的差值也越小，这一点也符合我们的认识。图 6（b）是对差值的分布进行统计，从这个统计可以看出，这一差值绝大多数都分布在 0 值左右，有部分存在这较大的偏差。导致存在这种差异有很多原因，其中包括提取样本的精度（2.1 节中对两种参考样本的分析），图像模拟中加入的信号噪声、背景噪声和设备的噪声及利用 SExtractor 对天体进行提取的参数也存在一定差异。





(a)有效半径之差随着天体亮度的分布　　　　(b)有效半径之差的统计分布

图 6 模拟图像所提取天体的有效半径分析图

Fig. 6 Anayasis of Effective Radius for Galaxies from Simulation Image

图 7 是对模拟图像所提取的天体的亮度与参考数据的天体亮度进行的对比，对比的方法同图 6(a) 一样，通过亮度之差随着天体亮度的分布进行分析。该图中，横坐标是亮度，纵坐标是参考数据中天体的亮度与相应的模拟数据中提取的天体的亮度之差。统计的样本依然是亮度小于 29 等天体。从该图中可以看出，这一亮度差值大部分分布在 0 轴的上下，偏差绝大多数都集中在 0.5 以下。从这张图中，还可以看出一个规律，对于亮度越大的天体，模拟后再提取，其亮度值越准确。

2.3 性能分析

本文中图像模拟软件开发以及统计的数据生成都是用 JAVA 程序开发。在天文中用 JAVA 开发的程序较少，本文之所以选择 JAVA 程序开发主要只因为 JAVA 程序存在着以下一些优势：1. JAVA 去掉了指针，有自动的垃圾回收机制，这样大大减小编写程序时由于内存泄露出现的问题，因此，能够将更多的时间用在算法的编写上。

2. JAVA 程序是面向对象的编程，在程序的可读性上要好于 C 程序。

3. JAVA 支持并行计算，且编写简单，对于本程序的并行计算上有很大的帮助。

另外，本文的程序运行在 Linux 系统下，硬件配置 CPU 为 Intel core i5 4 核 3.1GHz，内存为 16GB，在程序运行时，4 核全部使用，计算天体的数目为 4107，计算所需要的时间为 1267s，平均每个天体所需要时间为 308 ms。程序运行时峰值所需要的内存为 4.1GB。

程序运行的大部分时间和内存都是消耗在利用 Sersic 模型计算天体的形态，在计算天体的形态时，由于 1.4 节中所述子像素的需求以及计算的精度，需要将一个像素又分解成若干个小像素，计算的时间和内存的消耗和像素分解的个数成正比。

对于星系形态的计算采用了二维的数值积分，为了计算更加准确，本文将一个像元分解为若干个子像元，通过利用子像元积分得到每一个大的像元的准确数据，文中兼顾性能需求没有对像元进行无限划分，只是分解为 32×32 个子像素，4500×4500 个像素的图像就变成了 144,000×144,000 个像素的图像，这样数据量

已经很大，采用并行及算法上的优化使得每个天体的计算时间达到 308ms。

3 总结及展望

本文中对于图像模拟所用到的方法都是天文中比较基础的模型，例如 Sersic 模型、Gaussian 模型。将这些基础模型应用到模拟中，可以达到简单高效的目的。但是简单高效只是在模拟结果真实正确的基础上才能成立。通过本文第 3 部分的分析可以发现，模拟的结果与真实数据存在一定差异，但是基本特征都能够正确的表现出来（例如，大小，方向、亮度等）；通过统计分析可以看出，对于模拟图像提取出的天体的数目、亮度、大小与参考数据比较，错误的数量都在可以控制的范围内。另外，本文中对模拟程序实现了并行计算，而且在程序中也做了已做了一些优化（例如，三角函数用插值方法计算），通过 2 部分中对于性能分析可以看到，计算天体平均时间达到了毫秒级，再次说明了本文的模拟成像快速。

模拟程序中，点扩散函数只用到 Gaussian 函数，虽然简单，但是对于反映图像的真实性上还存在着一定差别。另外，模拟的目的主要是为一些新的巡天项目分析服务。所以本文的下阶段目标就是在模拟程序上对点扩散函数进行改进，加入真实光学系统特性；另外需要引入光谱的模拟以适应多波段成像模拟。同时，在分析上，将模拟与新的巡天项目相结合，得到与不同硬件相关的分析。

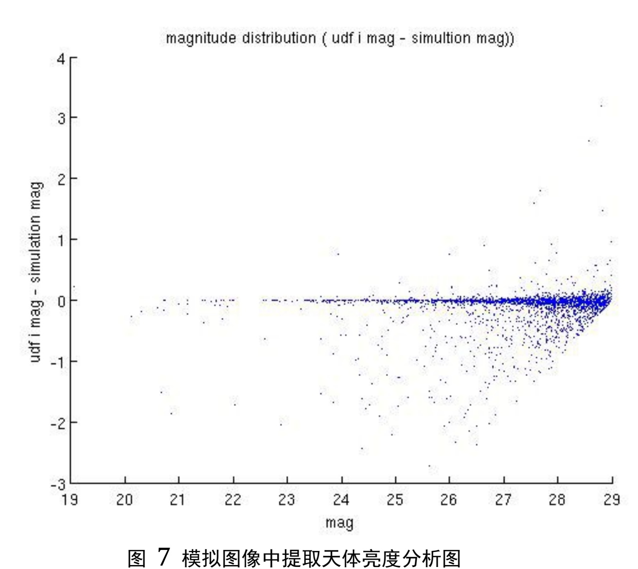


图 7 模拟图像中提取天体亮度分析图

Fig. 7 Magnitude analysis for Galaxies from Simulation Image

附录 1. SExtractor 提取主要参数

DETECT_MINAREA 3 # minimum number of pixels above threshold

DETECT_THRESH 1.2 # <sigmas> or <threshold>,<ZP> in mag.arcsec-2

ANALYSIS_THRESH 2 # <sigmas> or <threshold>,<ZP> in mag.arcsec-2

```
FILTER           Y              # apply filter for detection (Y or N)?
FILTER_NAME      default.conv   # name of the file containing the filter

DEBLEND_NTHRESH  32             # Number of deblending sub-thresholds
DEBLEND_MINCONT   0.005               # Minimum contrast parameter for
deblending
```

A fast image simulation method and analysis for astronomy research

Zhang Xin[1,2]

(1.Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Science, Beijing, 100101， China; 2. University of Chinese Academy of Science, Beijing, 100049, China)．

Abstract: Image simulation has played an increasingly important role in astronomical research. There are several important effects as follow: Through simulation, it can provide an important basis for the evaluation of astronomical observation equipment; Data processing algorithm can be verified through the processing of simulation data. In this paper, we describe the whole process of image simulation and optimize the local methods of image simulation, so that the simulation results can be generated efficiently and accurately. The main work of this paper includes the following aspects: Simulation for astronomical objects, including point sources (such as stars) and extended source (galaxies), and mainly introduce the galaxy simulation in this paper; Simulation of observation conditions, mainly include instrument Noise, stray light etc.; Simulation of optical system character which shows in image as point spread function (PSF); Optimization of calculation flow from calculations and procedural methods; Analysis for extracted galaxies sample. The data used in this paper is Hubble Ultra Deep Space (HUDF) ACS WFC i-band (F775) data with a limit of magnitude equal to 29 AB Mag.

Key words: Image Simulation; HUDF; Galaxy

参考文献:

1 ［1］Bertin E. SkyMaker: astronomical image simulations made easy[J]. Memorie della Societa Astronomica Italiana, 2009, 80:422.

2 ［2］Refregier A. Shapelets - I. A method for image analysis[J]. Monthly Notice of the Royal Astronomical Society, 2003, 338(1): 35-47.

3 ［3］LSST Science Collaborations, et al. LSST Science Book Version 2.0 [EB/OL].2009, arXiv:0912.0201.

4 ［4］Bergé J., Gamper, L., et al. An Ultra Fast Image Generator (UFig) for wide-field astronomy[J]. A&C, 2013, 1:23- 32.

5 ［5］Sérsic J. L. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy [J]. Boletin de la Asociacion Argentina de Astronomia , 1963, 6: 41.

6 ［6］Coe D., Benitez N. et al. Galaxies in the Hubble Ultra Deep Field: I.Detection, Multiband Photometry, Photometric Redshifts, and Morphology[J]. The Astronomical Journal , 2006, 132(2): 926-959.

7 ［7］Beckwith S. V. W., Stiavelli M., Koekemoer A. M., et al. The Hubble Ultra Deep Field [J]. The Astronomical Journal , 2006, 132(5):1729-1755.

8 ［8］Diaz, R. I.; McLean, D.; Greenfield, P. The HST Exposure Time Calculators: Estimating Accurate Observing Times for HST Observations[C]. San Francisco: Astronomical Society of the Pacific, 2010.

9 ［9］Moffat, A. F. J. A Theoretical Investigation of Focal Stellar Images in the Photographic Emulsion and Application to Photographic Photometry[J]. Astronomy and Astrophysics, 1969, 3:455.

10 ［10］Trujillo, I., Aguerri, J. A. L., Cepa, J., Gutierrez, C. M. The effects of seeing on Sersic profiles - II. The Moffat PSF[J]. Monthly Notices of the Royal Astronomical Society, 2001, 328(3): 977-985.